Replication Guide for 'Estimating the SARS-CoV-2 Infection Fatality Rate by Data Combination: The Case of Germany's First Wave' by Thomas Dimpfl, Jantje Sönksen, Ingo Bechmann and Joachim Grammig, The Econometrics Journal, 2022

Overview

The codes in this package replicate all figures and in-line computations in Dimpfl, Sönksen, Bechmann and Grammig (The Econometrics Journal, 2022). Four separate MATLAB programs are provided to be used for that purpose: `corona_4plots_public.m`, `corona_plot_public.m`, `pvals_Germany_public.m`, and `sensitivity_deathshift_public.m`. The data used and supplied in the replication package are public data provided by the Robert-Koch Institute (RKI) and they are combined with data provided by Streeck et al. (2020; Fig. 2c), and Knabl (2020; Table 1). A base MATLAB installation on a generic desktop machine under Windows or Linux is sufficient to perform a full replication analysis.

Data Availability and Provenance Statements

Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript. In particular, all data are in the public domain.

Summary of Availability

All data are publicly available.

Details on each Data Source

The source for the dataset "Fallzahlen_RKI.csv" is Robert Koch Institute (2020a), the source for the dataset "Nowcasting_Zahlen.csv" is Robert Koch Institute (2020b), the source for the dataset "testzahlen.csv" is Robert Koch Institute (2020c). The source for the data used to estimate the detection probabilities are Knabl et al (2020, Table 1) and Streeck (2021, Figure 2c). This input is directly coded in the MATLAB programs (see below).

Dataset list

Data file	Source	Notes	Provided
`Fallzahlen_RKI.csv`	Robert Koch Institute (2020a)	Contains information on the number of documented cases and deaths.	Yes
`Nowcasting_Zahlen.csv`	Robert Koch Institute (2020b)	Contains RKI nowcast of infections.	Yes
`testzahlen.csv`	Robert Koch Institute (2020c)	Contains information on the number of SARS-CoV-2 tests that were conducted in Germany (weekly basis).	Yes

Computational requirements

Software Requirements

The only software needed is MATLAB with no special toolboxes. The MATLAB programs were run with release 2021a on a generic laptop under Windows 10. We also tested the programs on a Linux installation (OpenSuse Leap 15.0).

Memory and Runtime Requirements

- Approximate time needed to reproduce the analyses on a standard (year 2022) desktop machine: <1 minute for each of the MATLAB programs.
- The codes were last run on a 4-core Intel-based laptop with Windows 10.

Description of programs/code

- The MATLAB program `corona_4plots_public.m` creates the graphs depicted in Figure 1 of the paper.
  The input datasets are "Fallzahlen_RKI.csv" (Robert Koch Institute (2020a)), "Nowcasting_Zahlen.csv" (RKI 2020b) and "testzahlen.csv" (Robert Koch Institute (2020c)).
  Four .eps files are created that contain the graphs for the four panels in Figure 1 of the paper, "test2.eps", "Nowcast2.eps", "Death2.eps", and "CFR2_Germany_deaths19.eps"
- The MATLAB program `corona_plot_public.m` creates the graphs depicted in Figure 2 of the paper and computes the infection fatality rate (IFR) point estimates and the confidence intervals reported in the paper.
  The input datasets are "Fallzahlen_RKI.csv" (Robert Koch Institute (2020a)), "Nowcasting_Zahlen.csv" (RKI 2020b). The data input from Knabl et al (2020, Table 1) and Streeck (2021, Figure 2c) is directly coded in the program (lines 75 to 81).
  To produce the results based on the Ischgl data, set "regn='Ischgl'" in line 9 of the code. To produce the results based on the Gangelt data, set "regn='Gangelt'" instead.
  The program writes the IFR point estimate (in %) and the bounds of the 95% confidence interval reported in the paper in text files named "IFR_May1_Germany_$study$_deaths19.txt", where $study$ refers to either 'Ischgl' or 'Gangelt', depending on the chosen value for the variable "regn".
  The results – point estimates and 95% confidence interval – are also printed in the output window.
  With "regn='Ischgl'" you should find printed in the MATLAB output window:

"This is the IFR point estimate (in %): 0.8337
This is the 95% confidence interval (in %): 0.68701 0.98039"
With "regn='Gangelt'" you should find printed in the MATLAB output window:
"This is the IFR point estimate (in %): 0.84903
This is the 95% confidence interval (in %): 0.52415 1.1739"
The program also produces the .eps files used for Figure 2 in the paper. They are named "IFR_Germany_$study$_deaths19_Zoom.eps", where $study$ is either 'Ischgl' or 'Gangelt', depending on the chosen value for the variable "regn".
The program also produces .eps files which are not used in the paper, and which show an extended range for which the IFR estimates are computed: "IFR_Germany_$study$_deaths19.eps", where $study$ is either 'Ischgl' or 'Gangelt', depending on the chosen value for the variable "regn".
The program saves the data generated for the plots (the IFR estimates) in Matlab data files "ifr_ser_Germany_$study$_death19.mat", where $study$ is either 'Ischgl' or 'Gangelt', depending on the chosen value for the variable "regn".

- The MATLAB program `pvals_Germany_public.m` tests for equality of the two IFRs (Gangelt- and Ischgl-based) and computes the associated p-value.
  The input datasets are "Fallzahlen_RKI.csv" (Robert Koch Institute (2020a)), "Nowcasting_Zahlen.csv" (RKI 2020b). The data input from Knabl et al (2020, Table 1) and Streeck (2021, Figure 2c) is directly coded in the program (lines 57 to 60).
  The result is printed in the output window which should read
  "This is the p-value on 01-05-2020: 0.93245"
- The MATLAB program `sensitivity_deathshift_public.m`, creates the graphs depicted in Figure 3 of the paper.
  The input datasets are "Fallzahlen_RKI.csv" (Robert Koch Institute (2020a)), "Nowcasting_Zahlen.csv" (RKI 2020b). The data input from Knabl et al (2020, Table 1) and Streeck (2021, Figure 2c) is directly coded in the program (lines 71 to 77).
  To produce the plot for Ischgl, set "regn='Ischgl'" in line 7 of the code. To obtain the plot using Gangelt data, set "regn='Gangelt'" instead.
  The program produces the .eps files used for Figure 3 in the paper. They are named "IFR_Germany_$study$_fatalityshift_01May2020.eps",

where $study$ is again either 'Ischgl' or 'Gangelt', depending on the chosen value for the variable "regn".

Instructions to Replicators

- The content of the replication zip-file should be extracted in a folder and then the four programs can be run in MATLAB.
- A standard desktop machine or laptop running under Windows or Linux is sufficient.
- The MATLAB codes were last run January 13, 2022 on a regular Windows 10 laptop.
- A base MATLAB installation is sufficient, no toolboxes are required. We ran the code under MATLAB release 2021a and 2020b, but older version will do.
- Run order does not matter.
- For each of the four programs, the replicator should expect the results in less than one minute.

List of tables and programs

The provided code reproduces:

- All numbers provided in text in the paper.
- All figures in the paper.

Figure/ Table #	Program	Output files	Note
Figure 1	`corona_4plots_public.m`	test2.eps, Nowcast2.eps, Death2.eps, CFR2_Germany_deaths19.eps	see program description above
Figure 2	`corona_plot_public.m`	IFR_May1_Germany _Gangelt_deaths19.txt, IFR_May1_Germany _Ischgl_deaths19.txt, IFR_Germany_Ischgl _deaths19_Zoom.eps, IFR_Germany_Gangelt _deaths19_Zoom.eps, IFR_Germany_Ischgl _deaths19.eps, IFR_Germany _Gangelt_deaths19.eps, ifr_ser_Germany _Gangelt_death19.mat, ifr_ser_Germany _Ischgl_death19.mat	see program description above
Figure 3	`sensitivity_deathshift_public.m`	IFR_Germany_Gangelt _fatalityshift_01May2020.eps, IFR_Germany_Ischgl _fatalityshift_01May2020.eps	see program description above
Table 1	n.a.	n.a.	Sources: Knabl et al. (2020; Table 1) and Streeck et al. (2020; Fig. 2c)
	`pvals_Germany_public.m`	Output is written in MATLAB output window	see program description above

References

Knabl, Ludwig and Mitra, Tanmay and Kimpel, Janine and Rössler, Annika and Volland, Andre and Walser, Andreas and Ulmer, Hanno and Pipperger, Lisa and Binder, Sebastian C and Riepler, Lydia and Bates, Katie and Bandyopadhyay, Arnab and Schips, Marta and Ranjan, Mrinalini and Falkensammer, Barbara and Borena, Wegene and Meyer-Hermann, Michael and von Laer, Dorothee. 2020. "High SARS-CoV-2 seroprevalence in children and adults in the Austrian ski resort Ischgl." 2020. Preprint. https://doi.org/10.1101/2020.08.20.20178533.

Robert Koch Institute. 2020a. Fallzahlen_RKI [dataset]. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Fallzahlen_Kum_Tab.html [Accessed Dec 12, 2020].

Robert Koch Institute. 2020b. Nowcasting_Zahlen [dataset]. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/Nowcasting.html [Accessed Dec 12, 2020].

Robert Koch Institute. 2020c. Testzahlen [dataset]. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html [Accessed Dec 12, 2020].

Streeck, Hendrik and Schulte, Bianca and Kümmerer, Beate and Richter, Enrico and Höller, Tobias and Fuhrmann, Christine and Bartok, Eva and Dolscheid-Pommerich, Ramona and Berger, Moritz and Wessendorf, Lukas and Eschbach-Bludau, Monika and Kellings, Angelika and Schwaiger, Astrid and Coenen, Martin and Hoffmann, Per and Nöthen, Markus and Eis-Hübinger, Anna-Maria and Exner, Martin and Schmithausen, Ricarda Maria and Schmid, Matthias and Hartmann, Gunther. 2020. "Infection fatality rate of SARS-CoV-2 in a super-spreading event in Germany." *Nature Communications*, 11. https://doi.org/10.1038/s41467-020-19509-y.